Urban-Rural Classification using Census data, 1851-1911.

Harry Smith and Robert J. Bennett

hjs57@cam.ac.uk rjb7@cam.ac.uk

Working Paper 6:
Working paper series from ESRC project ES/M010953:
Drivers of Entrepreneurship and Small Business

University of Cambridge, Department of Geography and Cambridge Group for the History of Population and Social Structure, Downing Place, Cambridge, CB2 3EN, UK.

November 2017

Comments are welcomed on this paper: contact the authors as above.

Keywords: Urban, Rural, Towns, Census

JEL Codes: N13, N83, N93, J21

Urban-Rural Classification using Census data, 1851-1911.

Harry Smith and Robert J. Bennett

Working Paper 6: ESRC project ES/M010953: Drivers of Entrepreneurship and Small Business, University of Cambridge.

1. Introduction

This paper develops an urban classification that can be used for the ESRC-supported project ES/M010953 *Drivers of Entrepreneurship and Small Businesses*. Its aim is to produce a classification based on the well-established Law-Robson urban classification but using the new data available from the e-version of the census. The classification will be available as codes in the entrepreneurs database deposited at UK Data Archive (UKDA). An overview of the project is provided in Working Paper 1, with a full list of other Working Papers given at the end of this paper.

The main source used by the ESRC project is transcripts of the census, mainly derived for the I-CeM electronic database for 1851-1911 produced by a team at the University of Essex, deposited at the UKDA: *The Integrated Census Microdata (I-CeM).*[1] This provides transcriptions of the original Census Enumerators Books (CEBs) as well as enhancing the data with various codes for household structure, relationships between people, and occupations.[2] Within the data is information on *employers* (those who employed others), *sole proprietor own account self-employed* (who employed no-one else), as well as employees, workers and the unoccupied. The information on employers and own account is the main

[1] K. Schürer, E. Higgs, A.M. Reid, E.M Garrett, *Integrated Census Microdata, 1851-1911, version V. 2 (I-CeM.2)*, (2016) [data collection]. UK Data Service, SN: 7481, http://dx.doi.org/10.5255/UKDA-SN-7481-1; enhanced; E. Higgs, C. Jones, K. Schürer and A. Wilkinson, *Integrated Census Microdata (I-CeM) Guide*, 2nd ed. (Colchester: Department of History, University of Essex, 2015).

[2] e.g. 'General Instruction', Census of England and Wales, Householder's Schedule, 1851.

subject of research in the *Drivers of Entrepreneurship and Small Businesses* project. This paper focuses on how the I-CeM data on geographical locations can be used to create urban classification codes that can be used for analysis of different employment status categories in different places.

The urban classification described here uses I-CeM at the parish and Registration Sub-District (RSD) level. It identifies which parishes in each census contained some part of a town. The definition of towns is a complex matter on which much ink has been spilled by historians, geographers and economists. A great deal of such discussion has concerned distinguishing small towns from large villages. Chris Law and Brian Robson started from the proposition that towns in the mid nineteenth century had populations greater than 2,500.[3] This definition was refined by reference to population density and contiguity of built-up space, but the basic population criteria of 2,500 remained fundamental to their analysis and no town with a population smaller than that appears in their data. Some scholars have viewed this population threshold as too rigid. They argue that it excluded smaller locations which were functionally towns. For example, in his discussion of small towns in Britain in the period 1840-1950, Stephen Royle adopted an upper population limit of 10,000 but set no lower limit, relying instead on the lists of towns created by Peter Clark and Jean Hosking for England, Harold Carter for Wales, and on the Census listings of towns for Scotland. In doing so he accepted the arguments of these sources that such locations functioned as, or were legally defined as towns whatever their size.[4]

Historians, therefore, have tended to define towns in two ways. First, by using population criteria (whether aggregate figures or densities). Secondly, by the presence of institutions that allowed a location to perform urban functions, for example, the existence of a market or certain organs of local government. There are other methods that could be used to distinguish between urban and rural locations, such as analysis of occupation structures, demographic indicators or the density and connectivity of transport links. However, the nature of the available data, particularly the absence of parish-level occupation statistics, has meant that historians have been reliant on population and function-based definitions. This situation has changed as new data sources have become available. The I-CeM data provide occupation information at all geographical levels. This allows new statistical approaches to the definition

[3] Law, 1967; Robson, 1973.
[4] Royle, 2000, 152.

and categorisation of towns to be undertaken. One such approach using factor analysis for the same towns as developed here is described in Working Paper 7 for the Entrepreneurship project.

While the new data open up important possibilities for alternative methods for defining towns, this Working Paper describes the creation of an urban/rural classification for England and Wales covering the period 1851-1911 and based primarily on population criteria. The classification is founded on the urban populations created by Chris Law and Brian Robson, although it interrogates those data and refines them. In doing so, this paper is the first close re-examination of the Law-Robson figures for the second half of the nineteenth century.[5]
It should be noted that in the current I-CeM project that for various reasons and despite all the efforts made, the allocation of individuals to parishes is not accurate for all parishes in all censuses.[6] Version 2 of the data being used here incudes a wide range of corrections introduced by Cambridge Group projects with the assistance of Kevin Schürer.[7] These data are also known still to contain some remaining errors of spatial attribution, but all allocations at RSD level were aligned to a level of at least 5% (in practice in almost all cases to less than 1%) of the published figures. At parish level the inaccuracies remaining relate to a few parishes that attribute a larger or smaller population compared to the GRO published record. In the urban classification developed here all parishes are scrutinised individually to determine their classification code based on their published populations. Consequently, the parishes allocated to towns are correct, even if the allocation of individuals to those parishes in I-CeM is incorrect. As the parish allocations are improved this classification will become more accurate. Furthermore, the misallocations in I-CeM are between parishes within the same RSD, consequently many of the misallocations in urban parishes are between parishes in the same town and thus do not affect the overall accuracy of the classification.

The rest of the paper is divided into three sections. Section 2 discusses the Law-Robson criteria used to classify towns. Section 3 discusses methods to identify urban parishes. Section 4 presents the urban classification. Section 5 provides the documentation of the urban classification data download.

[5] Jack Langton has examined and revised the Law-Robson data for the period 1801-41, see Langton, 2000.
[6] See Higgs et al., *Integrated Census Microdata (I-CeM) Guide*, 114-15.
[7] Schürer et al., *Integrated Census Microdata, 1851-1911, version V. 2*.

2. Law-Robson and criteria for inclusion as a town

The Law-Robson approach is concerned with identifying all towns over 2,500 population. For the purposes of this Working Paper the aim is to identify all large towns in each census year. The definition of a large town adopted is those towns with populations over 10,000 at the time of the Census. 10,000 is a commonly used population threshold. Adna Weber used this figure to distinguish between towns (locations with populations below 10,000) and cities (locations with populations above 10,000).[8] Jan de Vries also used the threshold of 10,000 to distinguish cities from towns in his study of European urbanization.[9] A population of 10,000 provides, therefore, a useful, albeit somewhat arbitrary, threshold to distinguish large towns and cities from smaller urban locations. It also provides a fairly natural cut-off of the size distribution for most years.

The urban population figures provided by Law and Robson were used to determine which places were included as large towns.[10] Law and Robson used four criteria to define a town and determine that town's population. First, the location's population had to be greater than 2,500. Secondly, the population density had to be greater than one person per acre. Thirdly, they used maps to identify the footprints of compact settlements or to include suburban areas within urban definitions. Fourthly, they used the degree of nucleation. The first two criteria were arbitrarily set to distinguish between the firmly urban and the rural and semi-rural, as Law commented, 'the figure of 2,500 was chosen because in practice it was found that this excluded the smaller market towns whose activities were very closely connected with the rural way of life'.[11] The second two criteria were used to mitigate the tendency for the population and density criteria to over-estimate urban populations. For example, the degree of nucleation was used to remove locations with high populations and densities but without any strong nucleus around which an urban way of life could develop; mining regions are particularly prone to this kind of settlement pattern. However, it is important to bear in mind the constraints that Law and Robson worked under, as they had only the published census tables available. Hence in practice, when some decisions were made, they were not always

[8] Weber, 1899, 16.
[9] de Vries, 2007, 22.
[10] The population data from Law and Robson, combined with those from Langton, are part of a UKDA deposit: R.J. Bennett, *Urban Population Database, 1801-1911* (2012). [data collection]. UK Data Service. SN. 7154, http://dx.doi.org/10.5255/UKDA-SN-7154-1; University of Salford, Department of Geography. University of Manchester, Department of Geography. University of Oxford. School of Geography and the Environment, [original data producers].
[11] Law, 1967, 129.

able accurately to attribute parishes. When Robson started with Law's data he confirmed as far as possible the Law allocations, but often had to work at a higher level of spatial aggregation: with local authority administrative areas such as Metropolitan Boroughs, County Boroughs, Sanitary Districts, etc. Both Law and Robson mapped parishes into this on a rather ad hoc basis depending on the published information available.[12]

For each census the towns selected based on their populations as given by Law and Robson are shown in Table 1.

Year	*Total*	*English towns*	*Welsh towns*
1851	152	145	7
1861	172	165	7
1871	215	202	13
1881	265	247	18
1891	311	288	23
1901	370	341	29
1911	406	371	35

Table 1. Towns over 10,000 in England and Wales, 1851-1911.
Source: Bennett (2012) *Urban Population Database, 1801-1911 (Law-Robson-Langton database).*

These figures are not as easily intelligible as they appear. Law and Robson aggregated several locations to anticipate future urban development. As Robson comments:

> Given the definitions adopted, had the peripheral boroughs or communities been regarded as discrete places before being submerged into the sprawl of their neighbouring giants, once they had been submerged they would have "died" as towns and the larger places correspondingly would have appeared to increase rather suddenly during the decade in which the amalgamation was assumed to have occurred. The largest urban areas have therefore *often* been amalgamated with certain of their neighbouring communities to

[12] We are grateful to Brian Robson for confirming the details of the allocation method used in practice; see also Bennett, 2012, *User Guide*.

> produce a generous definition and the population totals of the adjacent areas have been included in the overall total once the density criterion outlined above had been met.[13]

Thus, in Law and Robson Birmingham is combined with Smethwick, Liverpool with Birkenhead, Newcastle with Gateshead and so on. This is a reasonable methodology; however, it was not consistently applied by Law and Robson. For example, they combine the towns of Rochester, Gillingham, Strood and Chatham to form an urban unit they call the 'Medway Towns'. By 1911, Law and Robson's starting point, these four towns are indeed contiguous; however, other locations with similar settlement patterns remain disaggregated in the Law-Robson figures. For example, Hanley, Fenton, Burslem, Stoke-on-Trent, Tunstall and Longton, the six towns which make up the city of Stoke-on-Trent remain separate despite being as contiguous as the 'Medway Towns'. Similar issues affect the various settlements around the mouth of the Tyne; and Poole is separated from Bournemouth (which Law-Robson confusingly term Bournemouth and Poole: q.v. below). There are also difficulties associated with the South Wales valleys where the population is resident in a series of small connected centres which stretch the length of the valley. However, Law and Robson tend to lump all these settlements together and give them a generic name. For example, the Law-Robson town 'Rhondda' is not really a town; instead, it is the name of two separate valleys which are the site of a large number of settlements: Trehaford, Porth, Ynyshir, Wattstown, Tylorstown, Ferndale, Maerdy, Trealaw, Penygraig, Tonypandy, Clydach Vale, Llynypia, Ystrad, Pentre, Treorchy, Cwm Parc and Treherbert. None of these individual towns would have met the 10,000-population threshold. Indeed, the Rhondda valleys represent, for Britain, an unusual settlement pattern, contiguous linear settlement stretching over 14 miles down two valleys. It is reasonable to combine all these small individual settlements into one town because they represent an area of continuous high-density residence; however, the resulting 'town' is a different entity from the nucleated settlement of the majority of English and Welsh towns.

There are relatively few critical evaluations of Law and Robson's population figures. Table 2 reproduces Law's own comparison of his estimates with those of earlier scholars.

[13] Robson, 1973, 49, our emphasis.

Year	*Law-Robson*	*Census*	*Price-Williams*	*Weber*
1851	54.0	50.2	51.4	
1861	58.7	54.6	51.9	
1871	65.2	61.8	55.6	
1881	70.0	67.8		
1891	74.5	72.0		68.0
1901	78.0	77.0		
1911	78.9	78.1		

Table 2. Comparison of urban population estimates, percentage of population classified as urban. *Source*: Law, 'Urban population', 131, table VIII.

Certainly, the method adopted by Law and Robson was the most detailed study of the extent of urban areas and their populations. However, they were restricted by the form in which the population data was available. The data they used was available only from published census reports and aggregated at the parish level. For parishes which were partly urban and partly rural Law and Robson apportioned a proportion of their parish's population based on the extent to which the parish was urbanised using maps. However, it is not clear precisely how this was done and consequently recreating their figures with confidence is sometimes difficult. In general there is no record of the constituent parishes and part-parishes that Law-Robson used for their database. Hence to map and utilise their classification requires us to reconstruct their decisions, as far as this is now possible.

3. Method for identifying urban parishes

Several different approaches were tested before the final method described below was identified. First, and most basic, the population densities of all parishes and Registration Sub Districts (RSDs) were calculated and densities of either 1 person per acre (for parishes) or 0.3 persons per acre (for RSDs) were used to distinguish urban and rural locations. This approach was broad-brushed and provided no way of identifying parishes associated with particular towns. Furthermore, many towns cross parish and, even, RSD boundaries meaning that some form of aggregation was required.

Initially it was hoped that RSDs would provide a suitable balance between geographical specificity while involving a manageable number of units; for example, there are 2,110 RSDs in 1891 as opposed to 15,113 parishes. In this method, RSDs that contained part of a town would be aggregated. However, the complexity of the relationship between urban locations and the Census administrative geography meant that working out population aggregations by hand, even at RSD level, was too time-consuming and potentially inaccurate to be feasible. This highlighted the necessity of automating at least part of the process. The difficulty of matching RSDs to urban locations also suggested that it was necessary to work at the parish level. However, it was also clear that even at the parish level it would be difficult because many parishes were only partly urban. Until the census data can be broken down to a finer geographical level than the parish, any urban classification will be imperfect because of the necessity to include whole parish populations when in reality only part of any such population actually resided in a town.

Parish	*Population*
Bedford Eastern Ward St Paul	2,468
Bedford Western Ward St Paul	4,241
Bedford Eastern Ward St Cuthbert	663
Bedford Western Ward St Mary	1,334
Bedford Eastern Ward St Mary	336
Bedford Eastern Ward St John	56
Bedford Western Ward St John	397
Bedford Eastern Ward St Peter	2,109
Bedford Western Ward St Peter	89
Total	11,693

Table 3. Bedford parishes in 1851.
Source: I-CeM, 1851 data.

The method developed here is to use published parish populations totals combined with I-CeM RSD and parish data through a GIS comparison to attempt to match Law-Robson calculations of population size for the towns they identify for each census year 1851-1911. To use the Law-Robson population totals requires as a first step aggregation of the published parish populations associated with those locations in order to recreate the Law-Robson

population figures. For example, Law-Robson give the population of Bedford in 1851 as 11,693. This figure can be recreated by aggregating nine 1851 Census parishes (see Table 3).

Repeating this process for each town with a population over 10,000 in each census year would have taken a long time to do by hand. Consequently, a GIS file of an *urban footprint* created by researchers at the Cambridge Group for the History of Population and Social Structure was used to identify potential parishes within each urban area.[14] This seeks to parallel Law-Robson's use of maps and maintains a similar methodology, although it includes many smaller settlements as it does not rely on Law-Robson's population threshold of 2,500. This GIS file contains the urban footprint of 1,606 potential towns in England and Wales drawn from maps dating from the late 1890s. This GIS was overlain on parish maps of England and Wales and the degree of overlap between parish and footprint was used to identify parishes covered in some part by a town's built-up area at the end of the nineteenth century. The parishes identified in this manner were further refined by using parish population densities to help account for the change in the size of towns over the second half of the nineteenth century. The parishes associated with each town were then checked by hand to obtain a population for each town which was within five per cent of the Law-Robson population for that town. Table 4 shows the degree to which these populations matched with Law-Robson's population figures for each census year.

Difference (%)	*1851*	*1861*	*1871*	*1881*	*1891*	*1901*	*1911*
<-20	3	4	4	8	11	4	11
-20:-10	1	4	10	11	9	9	7
-10:-5	4	6	9	11	18	9	9
-5:0	35	27	50	49	61	67	57
0	73	77	75	82	79	231	272
0:5	22	36	35	53	70	68	39
5:10	5	6	7	15	14	5	5
10:20	6	5	10	9	14	1	2
>20	3	7	14	27	35	2	2

Table 4. Recreating Law-Robson 1851-1911.

Source: *I-CeM*, 1881-1911 data.

[14] See the acknowledgements for full details of the GIS files used; the footprint file is derived from L. Shaw-Taylor, A.E.M. Satchell and A. MacKenzie, *Built-up areas of potentially urban settlements in England and Wales, 1891-1912* (draft version 2016).

For each year, the majority of the populations calculated are within 5 per cent of the figures given by Law-Robson for the towns in question. Those which are more than 5 per cent larger or smaller than the Law-Robson figures can be explained by several factors. As mentioned above, Law and Robson, for parishes which were part urban and part rural, attributed a proportion of that parish's population to an urban location. As the process discussed here is meant to identify parishes as urban or not, the parishes have not been split and consequently the figures produced will either over- or under-estimate the population of towns when compared to the Law-Robson figures. Similarly, there are a number of parishes which contain more than one town. Law-Robson split the population of such parishes between the towns in question, an option unavailable to us as discussed above. For example, Coseley and Sedgley were, until 1903, contained within the same parish. As a consequence the population of that parish has to be assigned to both urban locations causing over-estimates in both cases (see Table 5).

Year	*Census Parish population*		*Law-Robson population*		*Difference (%)*	
	Coseley	*Sedgley*	*Coseley*	*Sedgley*	*Coseley*	*Sedgley*
1881	36,574	36,574	21,700	14,874	68.54	145.89
1891	36,860	36,860	21,899	14,961	68.32	146.37
1901	38,170	38,170	22,219	15,951	71.79	139.30
1911	22,834	16,527	22,834	16,527	0	0

Table 5. Coseley and Sedgley, population 1881-1911.
Source: *I-CeM*, 1881-1911 data.

Once again, this is not a serious issue as the aim is to classify parishes as urban or otherwise. However, it does have the effect of making the match with Law-Robson's figures look less accurate than they it is in reality. The main consequence of the various issues with parish boundaries is that in the database produced by this exercise some towns are combined which appear separately in Law and Robson's data.

There is a marked improvement in the degree of match between the recreated figures and those produced by Law and Robson towards the end of the period. This reflects the extensive reorganisation of parishes in the late nineteenth century so that parishes became better fits with urban locations. The Local Government Act of 1894 provided for the division of

parishes which were situated in more than one urban sanitary district or straddled the boundary of an urban and a rural sanitary district.[15] Consequently, the match between parishes and towns is much closer in 1901 and 1911, as the 1901 Census Report comments.[16] This change means that the Law-Robson figures are easier to recreate and there is less evidence of them needing to divide up parishes which were partly rural and partly urban in order to allocate a proportion of their population to a town.

Much of the discrepancy between Law-Robson's figures and those produced by the method outlined in this paper is caused by the difficulties associated with parish boundaries. However, at times it is difficult to understand how Law and Robson arrived at their figures. For example, Bradford is consistently difficult to recreate. For 1891 Law-Robson gave it a population of 265,728 while the method described above only produced a population of 223,604. There are similar differences in other census years. However they arrived at their figure, it appears that Law and Robson did not check it consistently against the map evidence as the parishes corresponding to the built-up environment of Bradford are too small to produce the population given by Law and Robson. Consequently, while the Law and Robson figures remain the most accurate urban population data available, there is some doubt as to how closely it is possible or desirable to match all their towns with the definitions developed here. Another puzzling issue is the treatment of Bournemouth and Poole. Poole appears twice in the Law-Robson dataset, once on its own and once combined with Bournemouth. It is impossible to recreate either population total; for example, in 1891 Poole Law-Robson give the population of Poole alone as 21,185 and Bournemouth & Poole as 66,454, no combination of the parishes that constitute Bournemouth and Poole can recreate either of these numbers. It is believed that Law and Robson actually had two separate towns, one for Bournemouth and Poole combined and one for Poole alone..[17] Consequently, here it has been decided to treat Bournemouth and Poole as separate settlements throughout and to calculate population totals disregarding the Law-Robson estimates.

This effort to construct urban populations from census parishes confirms that the urban population estimates produced by Law and Robson are the most accurate estimates currently

[15] *Local Government Act*, 1894, 56 & 57 Vict, c.73.
[16] *Census of England and Wales, 1901. General Report, with appendices*, *Parliamentary Papers* (1904), CVIII, 13.
[17] This was confirmed by personal communication with Brian Robson: the original Poole data were for Poole borough throughout. The Bournemouth and Poole figure for 1871 (5,906) was for Bournemouth Sanitary District only, but then Bournemouth and Poole boroughs were given as a joint settlement from 1881.

available to historians. However, the figures produced here are more transparent than those produced by Law and Robson because there has been no division of parishes. While Law and Robson's division of parish populations into urban and non-urban may have been effective, it is impossible to judge or recreate their actions because they are not explicitly recorded in their publications. Consequently, until a more sophisticated method is developed to sub-divide parishes, the method developed here provides a practical means to move forwards.

4. Coding parishes and urban classification 1851-1911

Having identified all the parishes associated with large towns for each census year it is then possible to code all parishes according to whether they were urban or not in each census year. A classification developed form the above methodology allows the following four urban-rural groups to be defined:

1 Urban: All those parishes which were associated with a large (Law-Robson) town with a population over 10,000, i.e. those extracted to create the population totals.

2. Urban transition: Non-urban parishes in an RSD containing at least one urban parish that has been allocated to the urban category; i.e. the rest of such RSDs.

3. Transition: Parishes in RSDs containing no urban parishes where the whole RSD had a population density of more than 0.3 persons per acre.

4. Rural: All other parishes, i.e. those in RSDs that had a density of less than 0.3.

Tables 6 and 7 shows the four classes of parishes in each group, the numbers and the proportions of each type of parish in each census year.

Year	*Urban*	*Urban Transition*	*Transition*	*Rural*
1851	1,249	958	4,441	9,747
1861	1,243	1,070	4,393	9,488
1871	1,351	1,137	4,266	9,274
1881	1,372	1,604	3,636	8,685
1891	1,466	1,738	3,314	8,595
1901	1,354	2,043	3,183	8,683
1911	1,178	2,092	3,383	8,278

Table 6. Urban classification: number of parishes 1851-1911.

Year	*Urban*	*Urban Transition*	*Transition*	*Rural*
1851	7.62	5.84	27.09	59.45
1861	7.68	6.61	27.13	58.89
1871	8.43	7.09	26.62	57.86
1881	8.97	10.49	23.77	56.78
1891	9.70	11.50	21.93	56.87
1901	8.87	13.39	20.85	56.59
1911	7.89	14.01	22.66	55.44

Table 7. Urban classification: proportion of parishes (%) 1851-1911

It is striking that there is relatively little change in the proportion of each type of parish across this period. The slight increase to 1891 in urban parishes represents urbanization; the subsequent decrease is explained by the rationalisation of urban parishes discussed above. However, this picture of stability is not found when the population living in each of these types of locations is considered.

Year	*Urban*	*Urban Transition*	*Transition*	*Rural*
1851	7,854,342	619,756	4,683,214	4,770,288
1861	9,714,419	735,695	5,021,308	4,594,802
1871	12,347,299	844,609	5,187,225	4,333,133
1881	15,662,788	1,427,435	4,807,183	4,077,032
1891	18,855,539	1,634,303	4,556,936	3,955,747
1901	22,479,988	2,141,442	4,130,096	3,776,317
1911	25,387,687	2,513,345	4,554,956	3,614,504

Table 8. Urban classification: population 1851-1911.

Year	*Urban*	*Urban Transition*	*Transition*	*Rural*
1851	48.81	3.46	26.12	26.60
1861	48.41	3.67	25.02	22.89
1871	54.36	3.72	22.84	19.08
1881	60.30	5.50	18.51	15.70
1891	65.01	5.64	15.71	13.64
1901	69.11	6.58	12.70	11.61
1911	70.38	6.97	12.63	10.02

Table 9. Urban classification: population (%) 1851-1911.

Tables 8 and 9 illustrate that this period witnessed rapid and sustained urbanization. Furthermore, the largest towns and cities (those with populations over 10,000) came to dominate the urban system, with nearly three quarters of the population living in these 406 large towns by 1911.

Figures 1-3 map the urban classification for the years 1851, 1891 and 1911. The data is mapped at the RSD level. In the method described above, if a RSD was not wholly constituted of urban parishes then it would contain both urban and urban transition parishes. Consequently, the urban parishes and urban transition parishes have to be combined and so the urban areas appear larger than they are in reality.

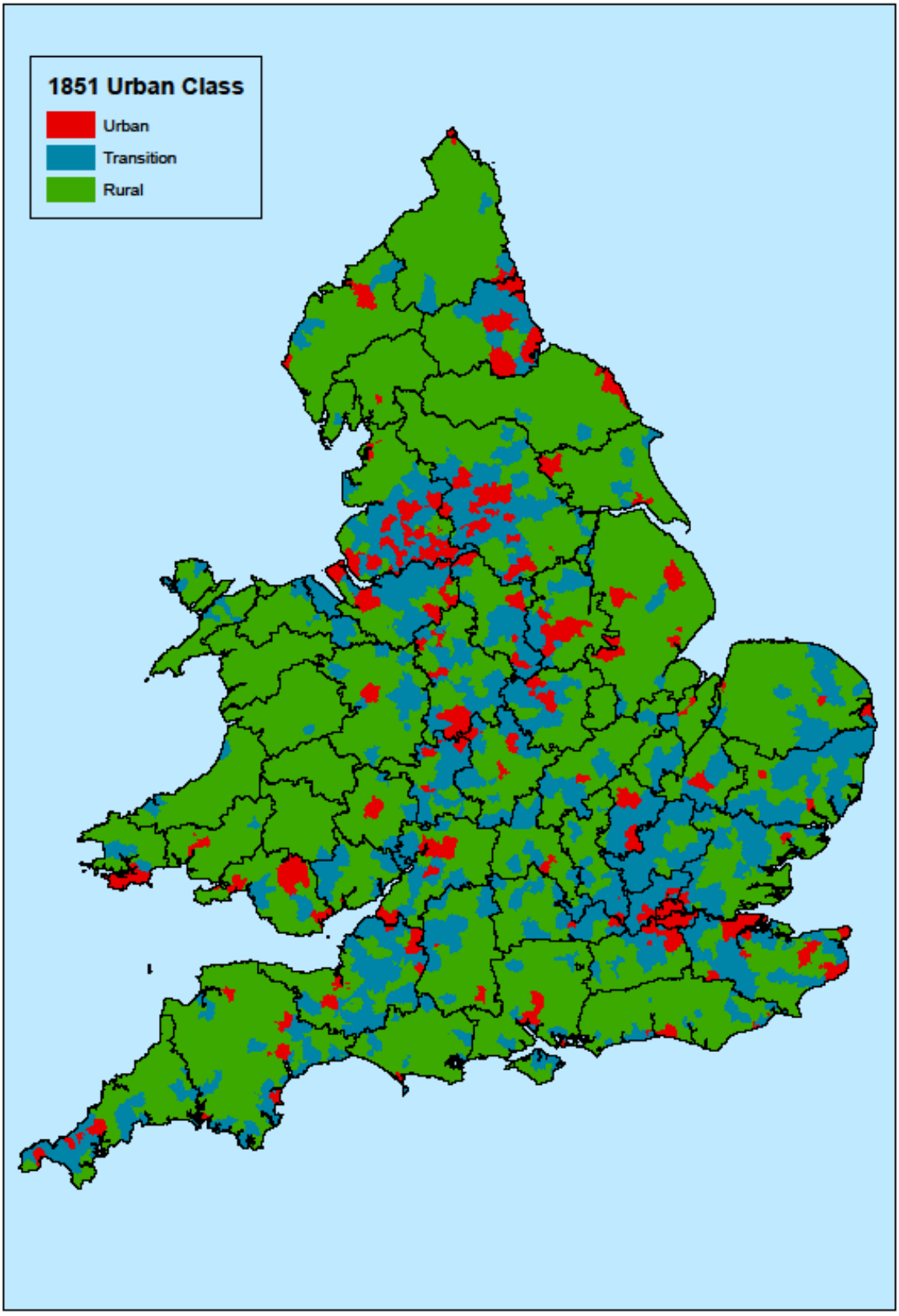

Figure 1. Urban classification, 1851.

ESRC project ES/M010953: WP 6: Smith et al.: *Urban-Rural Classification using Census data, 1851-1911*, Cambridge University.

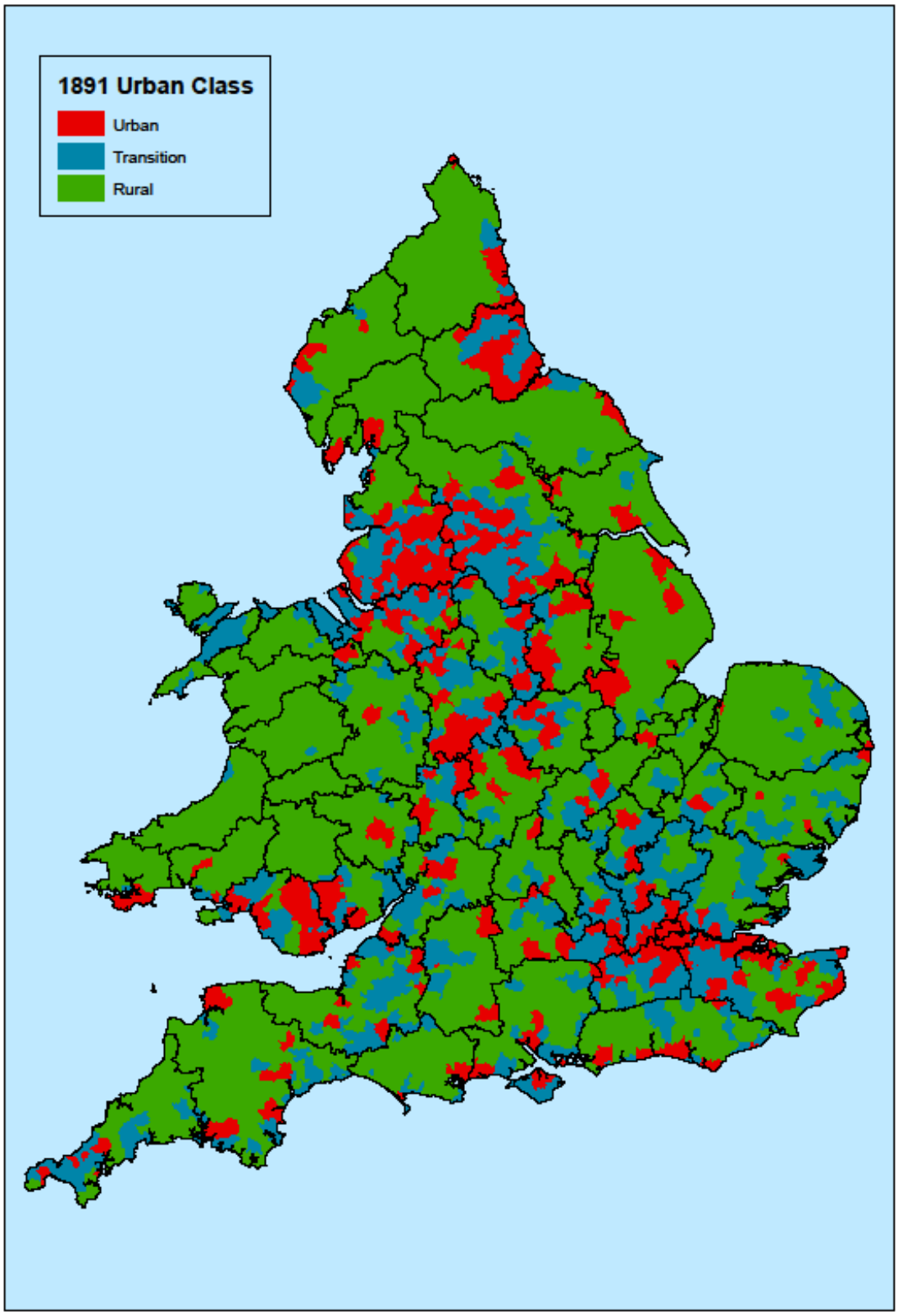

Figure 2. Urban classification, 1891.

ESRC project ES/M010953: WP 6: Smith et al.: *Urban-Rural Classification using Census data, 1851-1911*, Cambridge University.

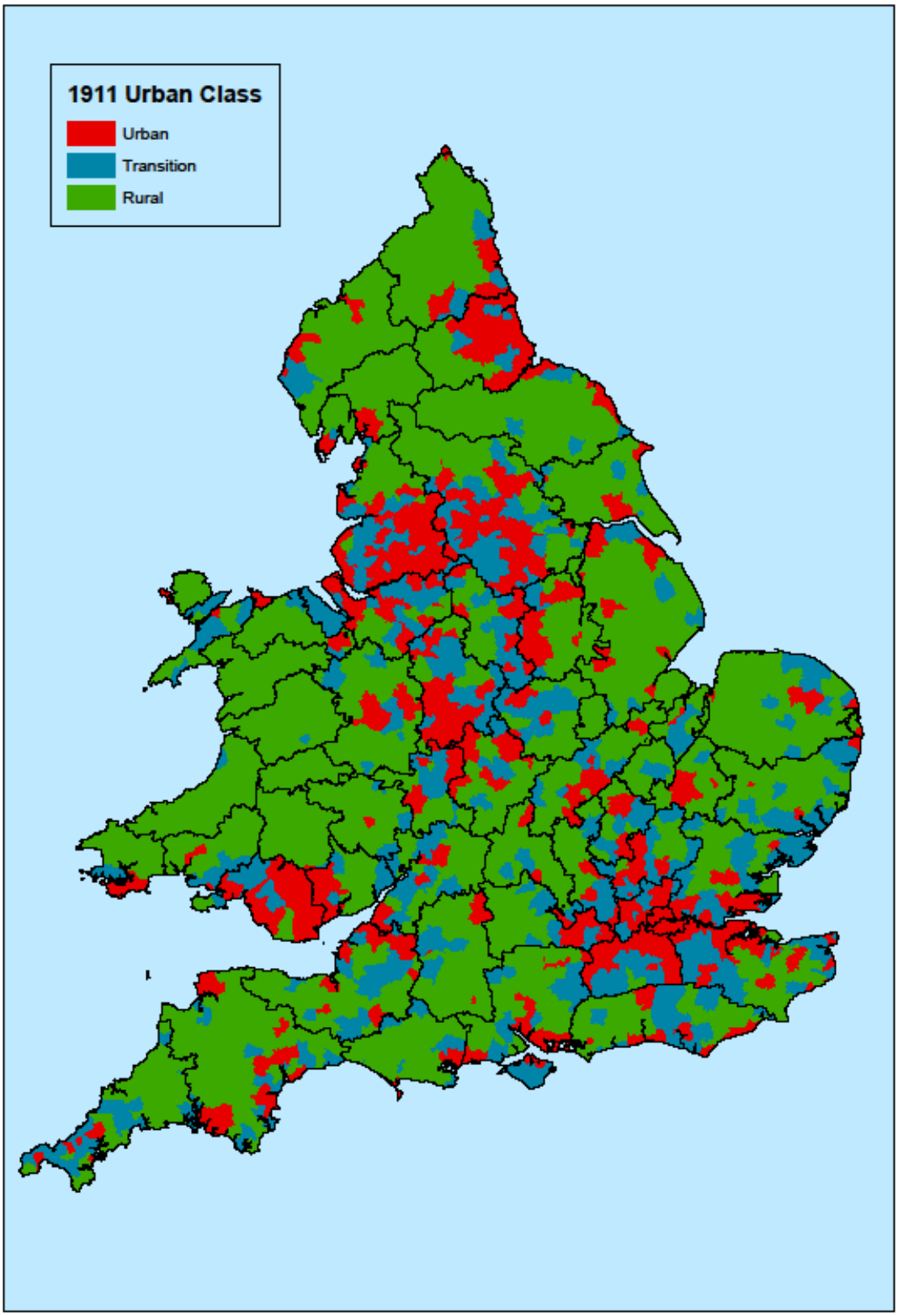

Figure 3. Urban classification, 1911.

The method adopted in this paper for identifying urban parishes also allows the reconstruction of urban populations because the parishes extracted as urban are each linked to a particular town. These towns have been coded using the town IDs created by Law and Robson. This allows detailed examination of different aspects of the characteristics of large towns. The occupational aggregation used is that developed in Working Paper 5.[18] For example, Table 10 gives the occupational structure of Birmingham in 1891.

Aggregate sector	*N*	*Aggregate sector*	*N*
Farming/estate work	2,714	Agricultural produce processing & dealing	1,786
Mining & quarrying	867	Food sales	10,334
Construction	15,878	Refreshment	2,784
Manufacturing	78,042	Finance & commerce	6,857
Maker-dealer	48,314	Public admin, military, clergy	2,429
Retail	5,392	Clerks, weighers, telegraph, non-theological students and apprentices	8,248
Transport	7,233	Domestic and service staff, cooks	10,100
Professional & business services	3,416	Labourers & transport staff (including family on farms)	12,817
Personal services	10,316	Persons of property and unoccupied	11

Table 10. Occupational structure, Birmingham, 1891.

Source: *I-CeM*, 1891 data.

This urban classification allows us to recreate the occupational classification of all large towns in Britain from 1851 onwards. Previous studies have been restricted in their analysis of urban occupational structure because of the necessity of relying either on the tabulations

[18] Bennett, Robert J., Smith Harry J., van Lieshout, Carry, and Newton, Gill (2017) *Business sectors, occupations and aggregations of census data 1851-1911.*

given in the published census reports or on painstaking recreation of individual settlements from the CEBs. The 1851 Census gave occupation statistics for counties, but not for any smaller geographic units. In 1861 and 1871 the data was given for registration districts and for the 'principal towns' in each division. The 1881 and 1891 census printed occupation breakdowns for every urban sanitary district with a population of more than 50,000. In 1901 the population criteria for inclusion dropped from 50,000 to 5,000, before returning to 50,000 in the 1911 Census. Occupational structure is just one aspect of urban settlements which can now be recreated; the urban classification allows analysis of a wide range of phenomena in an urban context: entrepreneurship, demographic data, migration, and so on. Furthermore, these issues can now be traced over time.

The ability to identify individual towns opens up a range of other possibilities for novel analysis. Figure 4 presents one simple example for entrepreneurship. This figure suggests there is no straightforward relationship between size of settlement and levels of entrepreneurship. Indeed, the five towns with the highest entrepreneurship rates in 1891 are not those that might immediately spring to mind when considering the topic of Victorian entrepreneurship. However, they are all resort towns, which suggests a relationship between leisure provision and high levels of business proprietorship and self-employment. The towns with the lowest entrepreneurship rates are also similar, four had large mining populations (Abertillery, Hebburn and Jarrow, Ashton-in-Makerfield, and Atherton), but a fifth, Crewe, was a centre for machine production and transport. This supports an argument in contemporary and historical scholarship that mining and other concentrated industries had a negative effect on entrepreneurship. This conclusion is assessed further in later research in this research project.

5. Documentation of the urban classification data download

This data file for the urban classification in its current form (which is still under development) may be made available for collaborative research by special request from the authors. It is final form it will be part of the full data deposit to be made at UKDA in 2018.

Variable	Description
CensusYear	Year of Census
ParID	I-CeM parish ID for that census year
UrbanClassification	Code for whether parish is urban, urban transition, transition or rural
TownID	Law-Robson town code
Town	Town name

Table 11. Urban classification data download structure.

6. Conclusion

This paper has presented the method used to create an urban classification for all censuses between 1851 and 1911. The method is based on recreating and developing the urban populations calculated by Chris Law and Brian Robson. The classification produced is based on large towns only, i.e. those with populations greater than 10,000 at the date of each census. The resulting classification produces two useful outputs. First, every parish in each census has a code which indicates whether it was urban, rural or transitional. Secondly, the parishes which constitute all large towns have been identified and coded. These two outputs allow the analysis of differences between urban and rural locations, and the analysis of differences within the urban system.

Acknowledgments:

This research has been supported by the ESRC under project grant ES/M010953: **Drivers of Entrepreneurship and Small Businesses.** Piloting of the research for 1881 draws from Leverhulme Trust grant RG66385: **The long-term evolution of Small and Medium-Sized Enterprises (SMEs).**

The population data from Law and Robson, combined with those from Langton, are derived from R.J. Bennett, *Urban Population Database, 1801-911* (2012). [data collection]. UKDA Data Service, SN: 7154, http://dx.doi.org/10.5255/UKDA-SN-7154-1; Chris Law, University of Salford, Department of Geography; Brian Robson, University of Manchester, Department of Geography; Jack Langton, University of Oxford, School of Geography and the Environment, [original data producers].

We are grateful for further information on the original classification from Chris Law and Brian Robson which has been used for some interpretations.

The census database used derives from K. Schürer, E. Higgs, A.M. Reid, E.M Garrett, *Integrated Census Microdata, 1851-1911, version V. 2 (I-CeM.2),* (2016) [data collection]. UK Data Service, SN: 7481, http://dx.doi.org/10.5255/UKDA-SN-7481-1; enhanced; E. Higgs, C. Jones, K. Schürer and A. Wilkinson, *Integrated Census Microdata (I-CeM) Guide*, 2nd ed. (Colchester: Department of History, University of Essex, 2015). A special acknowledgement of thanks is made to Kevin Schürer for advice and all his help in developing improved versions of I-CeM, and to Alice Reid, Eilidh Garrett, Joe Day, Hanna Jaadla, Xuesheng You, Leigh Shaw-Taylor and other members of the Campop I-CeM group who, with the authors, have collectively worked on the new versions of I-CeM.

The GIS boundary files for RSDs were constructed by Joe Day for the ESRC fertility project directed by Alice Reid:

http://www.geog.cam.ac.uk/research/projects/victorianfertilitydecline/publications.html

These used as a starting point the GIS parish files of Satchell, A.E.M., Kitson, P.M.K., Newton, G.H., Shaw-Taylor, L., Wrigley E.A. (2006) *1851 England and Wales census parishes, townships and places*, 2006, ESRC RES-000-23-1579, supported by Leverhulme Trust and the British Academy; Satchell, A.E.M. (2015) *England and Wales census parishes, townships and places*; which is an enhanced and corrected version of Burton, N, Westwood J., and Carter P. (2014) *GIS of the ancient parishes of England and Wales, 1500-1850*, UKDA, SN 4828; which is a GIS version of Kain, R.J.P., and Oliver, R.R. (2001) *Historic parishes of England and Wales: An electronic map of boundaries before 1850 with a gazetteer and metadata*, UKDA, SN 4348.

The urban footprint file was kindly provided by Leigh Shaw-Taylor and Max Satchell: Shaw-Taylor, L., Satchell, A.E.M., and MacKenzie, A., *Built-up areas of potentially urban settlements in England and Wales, 1891-1912* (draft version 2016), Cambridge Group for the History of Population and Social Structure.

References

Bennett, Robert J., (2012) *Urban Population Database, 1801-1911* (2012). [data collection]. UK Data Service. SN. 7154, http://dx.doi.org/10.5255/UKDA-SN-7154-1; University of Salford, Department of Geography. University of Manchester, Department of Geography. University of Oxford. School of Geography and the Environment, [original data producers].

de Vries, Jan (2007 [1984]), *European Urbanization, 1500-1800*, Abingdon: Routledge.

Higgs, E., C. Jones, K. Schürer and A. Wilkinson (2015) *Integrated Census Microdata (I-CeM) Guide*, 2nd ed., Department of History, University of Essex. UK Data Service, SN: 7481, http://dx.doi.org/10.5255/UKDA-SN-7481-1.

Langton, Jack (2000), 'Urban growth and economic change: from the late seventeenth century to 1841', in Peter Clark (ed.), *The Cambridge Urban History: Volume II: 1540-1840*, Cambridge: CUP, 453-90.

Law, Chris M. (1967), 'The growth of urban population in England and Wales, 1801-1911', *Transactions of the Institute of British Geographers*, 41, 125-43.

Robson, Brian (1973), *Urban Growth: An Approach*, London: Methuen.

Royle, Stephen A. (2000), 'The development of small towns in Britain', in Martin Daunton (ed.), *The Cambridge Urban History of Britain: Volume III: 1840-1950*, Cambridge: CUP, 151-84.

Schürer, K., Higgs, E., Reid, A. M. and Garrett, E. M. (2016) *Integrated Census Microdata, 1851-1911, version V. 2 (I-CeM.2),* [data collection].

Weber, Adna (1899), *The Growth of Cities in the Nineteenth Century: A Study in Statistics,* London: Macmillan.

Other Working Papers:

Working paper series: ESRC project ES/M010953: *'Drivers of Entrepreneurship and Small Business',* University of Cambridge, Department of Geography and Cambridge Group for the History of Population and Social Structure.

WP 1: Bennett, Robert J., Smith Harry J., van Lieshout, Carry, and Newton, Gill (2017) *Drivers of Entrepreneurship and Small Businesses: Project overview and database design.*

WP 2: Bennett, Robert J., Smith Harry J. and van Lieshout, Carry (2017) *Employers and the self-employed in the censuses 1851-1911: The census as a source for identifying entrepreneurs, business numbers and size distribution.*

WP 3: van Lieshout, Carry, Bennett, Robert J., Smith, Harry J. and Newton, Gill (2017) *Identifying businesses and entrepreneurs in the Censuses 1851-1881*.

WP 4: Smith, Harry J., Bennett, Robert J., and van Lieshout, Carry (2017) *Extracting entrepreneurs from the Censuses, 1891-1911*.

WP 5: Bennett, Robert J., Smith Harry J., van Lieshout, Carry, and Newton, Gill (2017) *Business sectors, occupations and aggregations of census data 1851-1911*.

WP 6: Smith, Harry J. and Bennett, Robert J. (2017) *Urban-Rural Classification using Census data, 1851-1911*.

WP 7: Smith, Harry, Bennett, Robert J., and Radicic, Dragana (2017) *Classification of towns in 1891 using factor analysis.*

WP 8: Bennett, Robert J., Smith, Harry, and Radicic, Dragana (2017) *Classification of occupations for economically active: Factor analysis of Registration Sub-Districts (RSDs) in 1891.*

Full list of all current Working Papers available at:
http://www.geog.cam.ac.uk/research/projects/historyofentrepreneurship/